# Nonlinear Spectral Unmixing of Hyperspectral Images Using Gaussian Processes

Yoann Altmann, *Student Member, IEEE*, Nicolas Dobigeon, *Member, IEEE*, Steve McLaughlin, *Fellow, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—This paper presents an unsupervised algorithm for nonlinear unmixing of hyperspectral images. The proposed model assumes that the pixel reflectances result from a nonlinear function of the abundance vectors associated with the pure spectral components. We assume that the spectral signatures of the pure components and the nonlinear function are unknown. The first step of the proposed method estimates the abundance vectors for all the image pixels using a Bayesian approach an a Gaussian process latent variable model for the nonlinear function (relating the abundance vectors to the observations). The endmembers are subsequently estimated using Gaussian process regression. The performance of the unmixing strategy is first evaluated on synthetic data. The proposed method provides accurate abundance and endmember estimations when compared to other linear and nonlinear unmixing strategies. An interesting property is its robustness to the absence of pure pixels in the image. The analysis of a real hyperspectral image shows results that are in good agreement with state of the art unmixing strategies and with a recent classification method.**

***Index Terms*—Gaussian processes, hyperspectral imaging, spectral unmixing.**

## I. Introduction

SPECTRAL UNMIXING (SU) is a major issue when analyzing hyperspectral images. It consists of identifying the macroscopic materials present in an hyperspectral image and quantifying the proportions of these materials in the image pixels. Many SU strategies assume that pixel reflectances are linear combinations of pure component spectra [1]–[4]. The resulting linear mixing model (LMM) has been widely adopted in the literature and has provided some interesting results. However, as discussed in [1], the LMM can be inappropriate for some hyperspectral images. Nonlinear mixing models provide an interesting alternative to overcome the inherent limitations of the LMM. For instance, the presence of relief can induce multiple scattering effects between the different materials present in the image. These nonlinear scattering effects typically occur in vegetation areas [5] and urban scenes [6]. A specific class of nonlinear models referred to as bilinear models has been studied in [5], [7]–[9] for modeling these multiple scattering effects. Conversely, the bidirectional reflectance-based model of [10] focusses on hyperspectral images including intimate mixtures, i.e., when the pure spectral components do not sit side-by-side in the pixel and when the photons are then interacting with all the materials simultaneously. These intimate mixtures can occur in sand-like or mineral areas. Other more flexible unmixing techniques have been also proposed to handle wider classes of nonlinearities, including radial basis function networks [11], [12] post-nonlinear mixing models [13] and kernel-based models [14]–[17].

Most existing unmixing strategies can be decomposed into two steps referred to as endmember extraction and abundance estimation. Endmember identification is usually achieved before estimating the abundances for all the image pixels. In the last decade, many endmember extraction algorithms (EEAs) have been developed to identify the pure spectral components contained in a hyperspectral image (see [1] for a recent review of these methods). Most EEAs rely on the LMM, which, as discussed, is inappropriate for the case of nonlinear mixtures of the endmembers. More recently, an EEA was proposed in [18] to extract endmembers from a set of nonlinearly mixed pixels. The nonparametric approach proposed in [18] was based on the assumption that the observed pixels lie on a manifold and that the endmembers are extreme points of this manifold. An approximation of the geodesic distance associated with this manifold was then used to identify the endmembers from the data. Similarly, our approach assumes that the data lie on a (possibly nonlinear) manifold. However, this paper proposes a parametrization of the manifold through the use of a kernel. The main advantages of this kernel-based parametric approach are 1) a better description of the manifold when the number of image pixels is reduced, and 2) a better endmember identification when there is no pure pixel in the image. This paper proposes first to estimate the abundance vectors and to estimate the endmembers during a second step, using the prediction capacity of Gaussian processes (GPs). This approach breaks from the usual paradigm of spectral unmixing. More precisely, this paper considers a kernel-based approach for nonlinear SU based on a nonlinear dimensionality reduction using a Gaussian process latent variable model (GPLVM). The main advantage of GPLVMs is their capacity to accurately model many different nonlinearities. In this paper, we propose to use a particular form of kernel based on existing bilinear models, which allows the proposed unmixing strategy to be accurate when the underlying mixing model is bilinear. Note that the LMM is a particular bilinear model. The algorithm proposed herein is "unsupervised" in the sense that

Y. Altmann, N. Dobigeon, and J.-Y. Tourneret are with the University of Toulouse, IRIT-ENSEEIHT, Toulouse 31071, France (e-mail: Yoann.Altmann@enseeiht.fr; Nicolas.dobigeon@enseeiht.fr; jean-yves.tourneret@enseeiht.fr).

S. McLaughlin is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: s.mclaughlin@hw.ac.uk).

Digital Object Identifier 10.1109/TSP.2013.2245127

the endmembers contained in the image and the mixing model are not known. Only the number of endmembers is assumed to be known. As a consequence, the parameters to be estimated are the kernel parameters, the endmember spectra and the abundances for all image pixels.

The paper is organized as follows. Section II presents the nonlinear mixing model considered in this paper for hyperspectral image unmixing. Section III introduces the GPLVM used for latent variable estimation. A constrained GPLVM for abundance estimation is detailed in Section IV. Section V studies the endmember estimation procedure using GP regression. Some simulation results conducted on synthetic and real data are shown and discussed in Sections VI and VII. Finally, conclusions are drawn in Section VIII.

## II. Nonlinear Mixing Model

Consider a hyperspectral image of $N$ pixels, composed of $R$ endmembers and observed in $L$ spectral bands. For convenience, the data are assumed to have been previously centered, i.e., the sample mean of the $N$ original pixels has been subtracted from each observed pixel. The $L$-spectrum $\mathbf{y}(n) = [y_1(n), \ldots, y_L(n)]^T$ of the $n$th mixed pixel ($n = 1, \ldots, N$) is defined as a transformation of its corresponding abundance vector $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ as follows

$$\mathbf{y}(n) = \boldsymbol{g}\left[\boldsymbol{a}(n)\right] + \mathbf{e}(n), \quad n = 1, \ldots, N \tag{1}$$

where $\boldsymbol{g} : \mathbb{R}^R \rightarrow \mathbb{R}^L$ is a linear or nonlinear unknown function. The noise vector $\mathbf{e}(n)$ is an independent, identically distributed (i.i.d.) white Gaussian noise sequence with variance $\sigma^2$, i.e., $\mathbf{e}(n) \sim \mathcal{N}(\mathbf{e}(n)|\mathbf{0}_L, \sigma^2\mathbf{I}_L)$, $n = 1, \ldots, N$. Without loss generality, the nonlinear mapping (1) from the abundance space to the observation space can be rewritten

$$\mathbf{y}(n) = \boldsymbol{W}_0\boldsymbol{\psi}\left[\boldsymbol{a}(n)\right] + \mathbf{e}(n), \quad n = 1, \ldots, N \tag{2}$$

where $\boldsymbol{\psi} : \mathbb{R}^R \rightarrow \mathbb{R}^D$, $\boldsymbol{W}_0$ is an $L \times D$ matrix and the dimension $D$ is the dimension of the subspace spanned by the transformed abundance vectors $\boldsymbol{\psi}[\boldsymbol{a}(n)]$, $n = 1, \ldots, N$. Of course, the performance of the unmixing strategy relies on the choice of the nonlinear function $\boldsymbol{\psi}$. In this paper, we will use the following nonlinearity

$$\begin{aligned} \boldsymbol{\psi} : \mathbb{R}^R &\rightarrow \mathbb{R}^D \\ \boldsymbol{a} &\mapsto \boldsymbol{\psi}[\boldsymbol{a}] = [a_1, \ldots, a_R, a_1a_2 \ldots, a_{R-1}a_R]^T, \end{aligned} \tag{3}$$

with $D = R(R+1)/2$. The primary motivation for considering this particular kind of nonlinearity is the fact that the resulting mixing model is a bilinear model with respect to each abundance $a_r, r = 1, \ldots, R$. More precisely, this mixing model extends the generalized bilinear model proposed in [9] and thus the LMM. It is important to note from (2) and (3) that $\boldsymbol{W}_0$ contains the $R$ spectra of the pure components present in the image and $R(R-1)/2$ interaction spectra between these components. Note also that the analysis presented in this paper could be applied to any other nonlinearity $\boldsymbol{\psi}$.

Due to physical constraints, the abundance vector $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ satisfies the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_r(n) = 1, \quad a_r(n) \geq 0, \forall r \in \{1, \ldots, R\}. \tag{4}$$

Since the nonlinearity $\boldsymbol{\psi}$ is fixed, the problem of unsupervised spectral unmixing is to determine the $L \times D$ spectrum matrix $\boldsymbol{W}_0$, the $N \times R$ abundance matrix $\mathbf{A} = [\boldsymbol{a}(1), \ldots, \boldsymbol{a}(N)]^T$ satisfying (2) under the constraints (4) and the noise variance $\sigma^2$. Unfortunately, it can be shown that the solution of this constrained problem is not unique. In the noise-free linear case, it is well known that the data are contained in a simplex whose vertices are the endmembers. When estimating the endmembers in the linear case, a simplex of minimum volume embedding the data is expected. Equivalently, the estimated abundance vectors are expected to occupy the largest volume in the simplex defined by (4). In a similar fashion to the linear case, the estimated abundance matrix resulting from an unsupervised nonlinear SU strategy may not occupy the largest volume in the simplex defined by (4). To tackle this problem, we first propose to relax the positivity constraints for the elements of the matrix $\mathbf{A}$ and to consider only the sum-to-one constraint. For ease of understanding, we introduce $R \times 1$ vectors satisfying the sum-to-one constraint

$$\sum_{r=1}^{R} x_r(n) = 1, \quad n = 1, \ldots, N \tag{5}$$

referred to as *latent variables* and denoted as $\boldsymbol{x}(n) = [x_1(n), \ldots, x_R(n)]^T$, $n = 1, \ldots, N$. The positivity constraint will be handled subsequently by a scaling procedure discussed in Section IV. The next section presents the Bayesian model for latent variable estimation using GPLVMs.

## III. Bayesian Model

GPLVMs [19] are powerful tools for probabilistic nonlinear dimensionality reduction that rewrite the nonlinear model (1) as a nonlinear mapping from a latent space to the observation space as follows

$$\mathbf{y}(n) = \boldsymbol{W}\boldsymbol{\psi}\left[\boldsymbol{x}(n)\right] + \mathbf{e}(n), \quad n = 1, \ldots, N \tag{6}$$

where $\boldsymbol{\psi}$ is defined in (3), $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L]^T$ is an $L \times D$ matrix with $\boldsymbol{w}_\ell = [w_{\ell,1}, \ldots, w_{\ell,D}]^T$, and $D = R(R+1)/2$. Note that from (2) and (6) the columns of $\boldsymbol{W}$ span the same subspace as the columns of $\boldsymbol{W}_0$. Consequently, the columns of $\boldsymbol{W}$ are linear combinations of the spectra of interest, i.e., the columns of $\boldsymbol{W}_0$. Note also that when $\boldsymbol{W}$ is full rank, it can be shown that the latent variables are necessarily linear combinations of the abundance vectors of interest. Figs. 1 and 2 illustrate the mapping from the abundance vectors to the observations that will be used in this paper. Note that the linear mapping between the abundances and the latent variables will be explained in detail in Section IV. For brevity, the $D \times 1$ vectors $\boldsymbol{\psi}[\boldsymbol{x}(n)]$ will be denoted as $\boldsymbol{\psi}_x(n)$ in the sequel. Assuming independence between the observations, the statistical properties of the noise

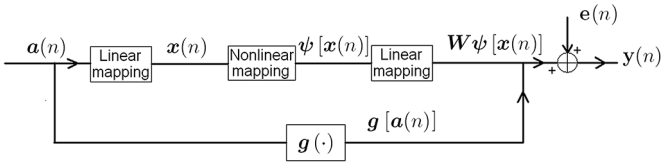



Fig. 1. Nonlinear mapping from the abundances vectors to the observed mixed pixels.

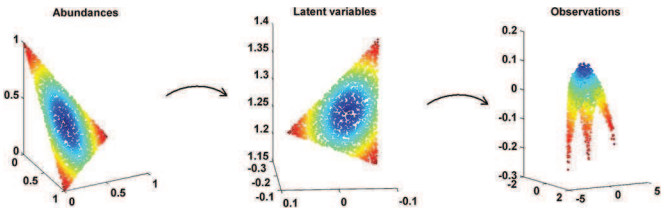



Fig. 2. Example of mapping decomposition from the abundance vectors to the observed nonlinearly mixed pixels through the latent variables ($R = 3$).

lead to the following likelihood of the $N \times L$ observation matrix $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)]^T$

$$\mathbf{Y}|\boldsymbol{W}, \boldsymbol{X}, \sigma^2 \sim \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{y}(n)|\boldsymbol{W}\boldsymbol{\psi}_x(n), \sigma^2\mathbf{I}_L\right) \tag{7}$$

where $\boldsymbol{X} = [\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)]^T$ is the $N \times R$ latent variable matrix. Note that the likelihood can be rewritten as a product of Gaussian distributions over the spectral bands as follows

$$\mathbf{Y}|\boldsymbol{W}, \boldsymbol{X}, \sigma^2 \sim \prod_{\ell=1}^{L} \mathcal{N}(\mathbf{y}_\ell|\boldsymbol{\Psi}_x\boldsymbol{w}_\ell, \sigma^2\mathbf{I}_L) \tag{8}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L]$ and $\boldsymbol{\Psi}_x = [\boldsymbol{\psi}_x(1), \ldots, \boldsymbol{\psi}_x(N)]^T$ is an $N \times D$ matrix. The idea of GPLVMs is to consider $\boldsymbol{W}$ as a nuisance parameter, to assign a Gaussian prior to $\boldsymbol{W}$ and to marginalize the joint likelihood (7) over $\boldsymbol{W}$, i.e.,

$$f(\mathbf{Y}|\boldsymbol{X}, \sigma^2) = \int f(\mathbf{Y}|\boldsymbol{W}, \boldsymbol{X}, \sigma^2) f(\boldsymbol{W}) \mathrm{d}\boldsymbol{W} \tag{9}$$

where $f(\boldsymbol{W})$ is the prior distribution of $\boldsymbol{W}$. The estimation of $\boldsymbol{X}$ and $\sigma^2$ can then be achieved by maximizing (9) following the maximum likelihood estimator (MLE) principle. An alternative consists of using an appropriate prior distribution $f(\boldsymbol{X}, \sigma^2)$, assuming prior independence between $\boldsymbol{W}$ and $(\boldsymbol{X}, \sigma^2)$, and maximizing the joint posterior distribution

$$\begin{aligned} f(\boldsymbol{X}, \sigma^2|\mathbf{Y}) &\propto \int f(\mathbf{Y}|\boldsymbol{W}, \boldsymbol{X}, \sigma^2) f(\boldsymbol{W}) f(\boldsymbol{X}, \sigma^2) \mathrm{d}\boldsymbol{W} \\ &\propto f(\boldsymbol{X}, \sigma^2) \int f(\mathbf{Y}|\boldsymbol{W}, \boldsymbol{X}, \sigma^2) f(\boldsymbol{W}) \mathrm{d}\boldsymbol{W} \\ &\propto f(\mathbf{Y}|\boldsymbol{X}, \sigma^2) f(\boldsymbol{X}, \sigma^2) \end{aligned} \tag{10}$$

with respect to (w.r.t.) $(\boldsymbol{X}, \sigma^2)$, yielding the maximum a posteriori (MAP) estimator of $(\boldsymbol{X}, \sigma^2)$. The next paragraph discusses different possibilities for marginalizing the joint likelihood (8) w.r.t. $\boldsymbol{W}$.

### A. Marginalizing $\boldsymbol{W}$

It can be seen from (9) that the marginalized likelihood and thus the associated latent variables depend on the choice of the prior $f(\boldsymbol{W})$. More precisely, assigning a given prior for $\boldsymbol{W}$ favors particular representations of the data, i.e., particular solutions for the latent variable matrix $\boldsymbol{X}$ maximizing the posterior (10). When using GPLVMs for dimensionality reduction, a classical choice [19] consists of assigning independent Gaussian priors for $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$, leading to

$$f(\boldsymbol{W}) = \left(\frac{1}{2\pi}\right)^{\frac{DL}{2}} \prod_{\ell=1}^{L} \exp\left(-\frac{1}{2}\|\boldsymbol{w}_\ell\|^2\right). \tag{11}$$

However, this choice can be inappropriate for SU. First, (11) can be incompatible with the admissible latent space, constrained by (5). Second, the prior (11) assumes the columns of $\boldsymbol{W}$ (linear combinations of the spectra of interest) are *a priori* Gaussian, which is not relevant for real spectra in most applications. A more sophisticated choice consists of considering *a priori* correlation between the columns (inter-spectra correlation) and rows (inter-bands correlation) of $\boldsymbol{W}$ using a structured covariance matrix to be fixed or estimated. In particular, introducing correlation between close spectral bands is of particular interest in hyperspectral imagery. Structured covariance matrices have already been considered in the GP literature for vector-valued kernels [20] (see [21] for a recent review). However, computing the resulting marginalized likelihood usually requires the estimation of the structured covariance matrix and the inversion of an $NL \times NL$ covariance matrix,[1] which is prohibitive for SU of hyperspectral images since several hundreds of spectral bands are usually considered when analyzing real data. Sparse approximation techniques might be used to reduce this computational complexity (see [23] for a recent review). However, to our knowledge, these techniques rely on the inversion of matrices bigger than $N \times N$ matrices. The next section presents an alternative that only requires the inversion of an $D \times D$ covariance matrix without any approximation.

### B. Subspace Identification

It can be seen from (6) that in the noise-free case, the data belong to a $D$-dimensional subspace that is spanned by the columns of $\boldsymbol{W}$. To reduce the computational complexity induced by the marginalization of the matrix $\boldsymbol{W}$ while considering correlations between spectral bands, we propose to marginalize a basis of the subspace spanned by $\boldsymbol{W}$ instead of $\boldsymbol{W}$ itself. More precisely, $\boldsymbol{W}$ can be decomposed as follows

$$\boldsymbol{W} = \boldsymbol{P}\boldsymbol{U}^T \tag{12}$$

where $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L]^T$ is an $L \times D$ matrix ($\boldsymbol{p}_\ell$ is $D \times 1$ vector) whose columns are arbitrary basis vectors of the $D$-dimensional subspace that contains the subspace spanned by the columns of $\boldsymbol{W}$ and $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_D]^T$ is a $D \times D$ matrix that scales the columns of $\boldsymbol{P}$. Note that the subspaces spanned by $\boldsymbol{P}$ and $\boldsymbol{W}$ are the same when $\boldsymbol{W}$ is full rank, resulting in a full rank matrix $\boldsymbol{U}$. The joint likelihood (8) can be rewritten as

$$\mathbf{Y}|\boldsymbol{P}, \boldsymbol{U}, \boldsymbol{X}, \sigma^2 \sim \prod_{\ell=1}^{L} \mathcal{N}(\mathbf{y}_\ell|\boldsymbol{C}\boldsymbol{p}_\ell, \sigma^2\mathbf{I}_L) \tag{13}$$

[1] See technical report [22] for further details.

where $\boldsymbol{C} = \boldsymbol{\Psi}_x \boldsymbol{U}$ is an $N \times D$ matrix. The proposed subspace estimation procedure consists of assigning an appropriate prior distribution to $\boldsymbol{P}$ (denoted as $f(\boldsymbol{P})$) and to marginalize $\boldsymbol{P}$ from the joint posterior of interest. It is easier to choose an informative prior distribution $f(\boldsymbol{P})$ that accounts for correlation between spectral bands than choosing an informative $f(\boldsymbol{W})$ since $\boldsymbol{P}$ is an arbitrary basis of the subspace spanned by $\boldsymbol{W}$, which can be easily estimated (as will be shown in the next section).

### C. Parameter Priors

GPLVMs construct a smooth mapping from the latent space to the observation space that preserves dissimilarities [24]. In the SU context, it means that pixels that are spectrally different have different latent variables and thus different abundance vectors. However, preserving local distances is also interesting: spectrally close pixels are expected to have similar abundance vectors and thus similar latent variables. Several approaches have been proposed to preserve similarities, including back-constraints [24], dynamical models [25] and locally linear embedding (LLE) [26]. In this paper, we use LLE to assign an appropriate prior to $\boldsymbol{X}$. First, the $K$ nearest neighbors $\{\mathbf{y}(j)\}_{j\in\nu_i}$ of each observation vector $\mathbf{y}(i)$ are computed using the Euclidian distance ($\nu_i$ is the set of integers $j$ such that $\mathbf{y}(j)$ is a neighbor of $\mathbf{y}(i)$). The weight matrix $\boldsymbol{\Lambda}_{\text{LLE}} = [\lambda_{i,j}]$ of size $N \times N$ providing the best reconstruction of $\mathbf{y}(i)$ from its neighbors is then estimated as

$$\boldsymbol{\Lambda}_{\text{LLE}} = \arg\min_{\boldsymbol{\Lambda}} \sum_{i=1}^{N} \left\| \mathbf{y}(i) - \sum_{j\in\nu_i} \lambda_{i,j}\mathbf{y}(j) \right\|^2. \quad (14)$$

Note that the solution of (14) is easy to obtain in closed form since the criterion to optimize is a quadratic function of $\boldsymbol{\Lambda}$. Note also that the matrix $\boldsymbol{\Lambda}$ is sparse since each pixel is only described by its $K$ nearest neighbors. The locally linear patches obtained by the LLE can then be used to set the following prior for the latent variable matrix

$$f(\boldsymbol{X}|\boldsymbol{\Lambda}_{\text{LLE}},\gamma) \propto \exp\left[-\frac{\gamma}{2}\sum_{i=1}^{N}\left\|\boldsymbol{x}(i) - \sum_{j\in\nu_i}\lambda_{i,j}\boldsymbol{x}(j)\right\|^2\right] \times \prod_{n=1}^{N} \mathbf{1}_{\mathcal{D}}\left[\boldsymbol{x}(n)\right] \quad (15)$$

where $\gamma$ is a hyperparameter to be adjusted and $\mathbf{1}_{\mathcal{D}}(\cdot)$ is the indicator function over the set $\mathcal{D}$ defined by the constraints (5).

In this paper, we propose to assign a prior to $\boldsymbol{P}$ using the standard principal component analysis (PCA) (note again that the data have been centered). Assuming prior independence between $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L$, the following prior is considered for $\boldsymbol{P}$

$$f(\boldsymbol{P}|\overline{\boldsymbol{P}}, s^2) = \left(\frac{1}{2\pi s^2}\right)^{\frac{NL}{2}} \prod_{\ell=1}^{L} \exp\left[-\frac{1}{2s^2}\|\boldsymbol{p}_\ell - \boldsymbol{p}_\ell\|^2\right] \quad (16)$$

where $\overline{\boldsymbol{P}} = [\bar{\boldsymbol{p}}_1, \ldots, \bar{\boldsymbol{p}}_L]^T$ is an $L \times D$ projection matrix containing the first $D$ eigenvectors of the sample covariance matrix of the observations (provided by PCA) and $s^2$ is a dispersion parameter that controls the dispersion of the prior. Note that the correlation between spectral bands is implicitly introduced through $\overline{\boldsymbol{P}}$.

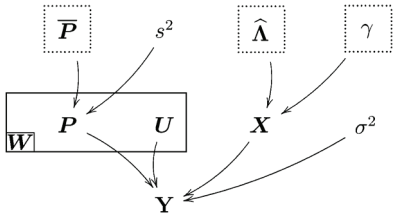



Fig. 3. DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

Non-informative priors are assigned to the noise variance $\sigma^2$ and the matrix $\boldsymbol{U}$, i.e,

$$\begin{aligned} f(\sigma^2) &\propto \mathbf{1}_{(0,\delta_{\sigma^2})}(\sigma^2) \\ f(u_{i,j}) &\propto \mathbf{1}_{(-\delta_{\boldsymbol{U}},\delta_{\boldsymbol{U}})}(u_{i,j}) \end{aligned} \quad (17)$$

where the intervals $(0, \delta_{\sigma^2})$ and $(-\delta_{\boldsymbol{U}}, \delta_{\boldsymbol{U}})$ cover the possible values of the parameters $\sigma^2$ and $\boldsymbol{U}$. Similarly, the following non-informative prior is assigned to the hyperparameter $s^2$

$$f(s^2) \propto \mathbf{1}_{(0,\delta_{s^2})}(s^2) \quad (18)$$

where the interval $(0, \delta_{s^2})$ covers the possible values of the hyperparameter $s^2$. The resulting directed acyclic graph (DAG) is depicted in Fig. 3.

### D. Marginalized Posterior Distribution

Assuming prior independence between $\boldsymbol{P}$, $\boldsymbol{X}$, $\boldsymbol{U}$, $s^2$ and $\sigma^2$, the marginalized posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{X}, \boldsymbol{U}, s^2, \sigma^2)$ can be expressed as

$$\begin{aligned} &f(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}_{\text{LLE}}, \overline{\boldsymbol{P}}, \gamma) \\ &\quad \propto f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) \int f(\mathbf{Y}|\boldsymbol{P}, \boldsymbol{\theta}) f(\boldsymbol{P}|\overline{\boldsymbol{P}}, s^2) \mathrm{d}\boldsymbol{P} \\ &\quad \propto f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\boldsymbol{P}}) f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) \end{aligned} \quad (19)$$

where $f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) = f(\boldsymbol{X}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) f(\boldsymbol{U}) f(s^2) f(\sigma^2)$. Straightforward computations leads to

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\boldsymbol{P}}) &= \int f(\mathbf{Y}|\boldsymbol{P}, \boldsymbol{\theta}) f(\boldsymbol{P}|\overline{\boldsymbol{P}}, s^2) \mathrm{d}\boldsymbol{P} \\ &\propto \prod_{\ell=1}^{L} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\bar{\boldsymbol{y}}_\ell^T \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{y}}_\ell\right] \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}}\bar{\mathbf{Y}}^T)\right] \end{aligned} \quad (20)$$

where $\boldsymbol{\Sigma} = s^2\boldsymbol{C}\boldsymbol{C}^T + \sigma^2\mathbf{I}_N$, $\bar{\boldsymbol{y}}_\ell = \mathbf{y}_\ell - \boldsymbol{C}\bar{\boldsymbol{p}}_\ell$ is an $N \times 1$ vector, $\bar{\mathbf{Y}} = [\bar{\boldsymbol{y}}_1, \ldots, \bar{\boldsymbol{y}}_L] = \mathbf{Y} - \boldsymbol{C}\overline{\boldsymbol{P}}^T$ is an $N \times L$ matrix and $\mathrm{tr}(\cdot)$ denotes the matrix trace.

Mainly due to the nonlinearity introduced through the non-linear mapping, a closed form expression for the parameters maximizing the joint posterior distribution (19) is impossible to obtain. We propose to use a scaled conjugate gradient (SCG) method to maximize the marginalized log-posterior. To ensure the sum-to-one constraint for $\boldsymbol{X}$, the following arbitrary reparametrization

$$x_R(n) = 1 - \sum_{r=1}^{R-1} x_r(n), \quad n = 1, \ldots, N$$

is used and the marginalized posterior distribution is optimized w.r.t. the first $(R-1)$ columns of $\boldsymbol{X}$ denoted $\boldsymbol{X}_{\setminus R}$. The partial derivatives of the log-posterior w.r.t. $\boldsymbol{X}_{\setminus R}, \boldsymbol{U}, s^2$ and $\sigma^2$ are obtained using partial derivatives w.r.t. $\boldsymbol{\Sigma}$ and $\bar{\mathbf{Y}}$ and the classical chain rules (see technical report [22] for further details). The resulting latent variable estimation procedure is referred to as locally linear GPLVM (LL-GPLVM).

Note that the marginalized likelihood reduces to the product of $L$ independent Gaussian probability density functions since

$$\mathbf{y}_\ell|\bar{\boldsymbol{p}}_\ell, \boldsymbol{U}, \boldsymbol{X}, \sigma^2, s^2 \sim \mathcal{N}(\boldsymbol{C}\bar{\boldsymbol{p}}_\ell, s^2\boldsymbol{C}\boldsymbol{C}^T + \sigma^2\mathbf{I}_N) \tag{21}$$

and $\ell = 1, \ldots, L$. Note also that the covariance matrix $\boldsymbol{\Sigma} = s^2\boldsymbol{C}\boldsymbol{C}^T\sigma^2\mathbf{I}_N$ is related to the covariance matrix of the 2nd order polynomial kernel [27, p. 89]. More precisely, the proposed nonlinear mapping corresponds to a particular polynomial kernel whose metric is induced by the matrix $\boldsymbol{U}$. Finally, note that the evaluation of the marginalized likelihood (20) only requires the inversion of the $N \times N$ covariance matrix $\boldsymbol{\Sigma}$. It can been seen from the following Woodbury matrix identity [28]

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\left[\mathbf{I}_N - \boldsymbol{C}(\sigma^2 s^{-2}\mathbf{I}_D + \boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\right] \tag{22}$$

that the computation of $\boldsymbol{\Sigma}^{-1}$ mainly relies on the inversion of a $D \times D$ matrix. Similarly, the computation of $|\boldsymbol{\Sigma}| = 1/|\boldsymbol{\Sigma}^{-1}|$ mainly consists of computing the determinant of a $D \times D$ matrix, which reduces the computational cost when compared to the structured covariance matrix based approach presented in Section III-A.

### *E. Estimation of $\boldsymbol{P}$*

Let us denote as $\hat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{U}}, \hat{s}^2, \hat{\sigma}^2)$ the maximum a posteriori (MAP) estimator of $\boldsymbol{\theta} = (\boldsymbol{X}, \boldsymbol{U}, s^2, \sigma^2)$ obtained by maximizing (19). Using the likelihood (13), the prior distribution (16) and Bayes' rule, we obtain the posterior distribution of $\boldsymbol{P}$ conditioned upon $\boldsymbol{\theta}$, i.e.,

$$\boldsymbol{P}|\mathbf{Y}, \boldsymbol{\theta}, \overline{\boldsymbol{P}} \sim \prod_{\ell=1}^{L} \mathcal{N}(\boldsymbol{p}_\ell|\hat{\boldsymbol{p}}_\ell\boldsymbol{S}) \tag{23}$$

where $\boldsymbol{S}^{-1} = \sigma^{-2}\boldsymbol{C}^T\boldsymbol{C} + s^{-2}\mathbf{I}_D$ and $\hat{\boldsymbol{p}}_\ell = \boldsymbol{S}(\boldsymbol{C}^T\mathbf{y}_\ell - \bar{\boldsymbol{p}}_\ell)$. Since the conditional posterior distribution of $\boldsymbol{P}$ is the product of $L$ independent Gaussian distributions, the MAP estimator of $\boldsymbol{P}$ conditioned upon $\hat{\boldsymbol{\theta}}$ is given by

$$\widehat{\boldsymbol{P}} = (\mathbf{Y}^T\widehat{\boldsymbol{C}} - \overline{\boldsymbol{P}})\widehat{\boldsymbol{S}} \tag{24}$$

where $\widehat{\boldsymbol{S}}^{-1} = \hat{\sigma}^{-2}\widehat{\boldsymbol{C}}^T\widehat{\boldsymbol{C}} + \hat{s}^{-2}\mathbf{I}_D$, $\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{\Psi}}_x\widehat{\boldsymbol{U}}$, $\widehat{\boldsymbol{\Psi}}_x = [\boldsymbol{\psi}_{\hat{x}}(1), \ldots, \boldsymbol{\psi}_{\hat{x}}(N)]^T$ and $\widehat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}(1), \ldots, \hat{\boldsymbol{x}}(N)]^T$. The next section studies a scaling procedure that estimates the abundance matrix using the estimated latent variables resulting from the maximization of (19).

## IV. Scaling Procedure

The optimization procedure presented in Section III-D provides a set of latent variables that represent the data but can differ from the abundance vectors of interest. Consider $\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{X}}_{\setminus R}, \mathbf{1}_N - \widehat{\boldsymbol{X}}_{\setminus R}\mathbf{1}_{R-1}]$ obtained after maximization of the posterior (19). The purpose of this section is to estimate an $N \times R$ abundance matrix $\mathbf{A} = [\boldsymbol{a}(1), \ldots, \boldsymbol{a}(N)]^T$ such that

$$\widehat{\boldsymbol{X}}_{\setminus R} = \mathbf{A}\boldsymbol{V}_{R-1}^T + \boldsymbol{E} \tag{25}$$

where $\boldsymbol{a}(1), \ldots, \boldsymbol{a}(N)$ occupy the maximal volume in the simplex defined by (4), $\boldsymbol{V}_{R-1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_R]$ is an $(R-1) \times R$ matrix and $\boldsymbol{E}$ is an $N \times (R-1)$ standard i.i.d Gaussian noise matrix which models the scaling errors. Since the rows of $\widehat{\boldsymbol{X}}$ satisfy the sum-to-one constraint (5), estimating the relation between $\widehat{\boldsymbol{X}}_{\setminus R}$ and $\mathbf{A}$ is equivalent to estimate the relation between $\widehat{\boldsymbol{X}}$ and $\mathbf{A}$. However, when considering the mapping between $\widehat{\boldsymbol{X}}$ and $\mathbf{A}$, non-isotropic noise has to be considered since the rows of $\widehat{\boldsymbol{X}}$ and $\mathbf{A}$ satisfy the sum-to-one constraint, i.e., they belong to the same $(R-1)$-dimensional subspace.

Eq. (25) corresponds to an LMM whose noisy observations are the rows of $\widehat{\boldsymbol{X}}_{\setminus R}$. Since $\mathbf{A}$ is assumed to occupy the largest volume in the simplex defined by (4), the columns of $\boldsymbol{V}_{R-1}$ are the vertices of the simplex of minimum volume that contains $\widehat{\boldsymbol{X}}_{\setminus R}$. As a consequence, it seems reasonable to use a linear unmixing strategy for the set of vectors $\widehat{\boldsymbol{x}}_{\setminus R}(1), \ldots, \hat{\boldsymbol{x}}_{\setminus R}(N)$ to estimate $\mathbf{A}$ and $\boldsymbol{V}_{R-1}$. In this paper, we propose to estimate jointly $\mathbf{A}$ and $\boldsymbol{V}_{R-1}$ using the Bayesian algorithm presented in [29] for unsupervised SU assuming the LMM. Note that the algorithm in [29] assumed positivity constraints for the estimated endmembers. Since these constraints for $\boldsymbol{V}_{R-1}$ are unjustified, the original algorithm has slightly been modified by removing the truncations in the projected endmember priors (see [29] for details). Once the estimator $(\widehat{\mathbf{A}}, \widehat{\boldsymbol{V}}_{R-1})$ of $(\mathbf{A}, \boldsymbol{V}_{R-1})$ has been obtained by the proposed scaling procedure, the resulting constrained latent variables denoted as $\widehat{\boldsymbol{X}}^{(c)} = [\hat{\boldsymbol{x}}^{(c)}(1), \ldots, \hat{\boldsymbol{x}}^{(c)}(N)]^T$ are defined as follows

$$\widehat{\boldsymbol{X}}^{(c)} = \left[\widehat{\boldsymbol{X}}_{\setminus R}^{(c)}, \quad \mathbf{1}_N - \widehat{\boldsymbol{X}}_{\setminus R}^{(c)}\mathbf{1}_{R-1}\right] \tag{26}$$

with $\widehat{\boldsymbol{X}}_{\setminus R}^{(c)} = \widehat{\mathbf{A}}\widehat{\boldsymbol{V}}_{R-1}^T$. Using the sum-to-one constraint $\widehat{\mathbf{A}}\mathbf{1}_R = \mathbf{1}_N$, we obtain

$$\begin{aligned}\widehat{\boldsymbol{X}}^{(c)} &= \left[\widehat{\mathbf{A}}\widehat{\boldsymbol{V}}_{R-1}^T, \quad \widehat{\mathbf{A}}\mathbf{1}_R - \widehat{\mathbf{A}}\widehat{\boldsymbol{V}}_{R-1}^T\mathbf{1}_{R-1}^T\right] \\ &= \widehat{\mathbf{A}}\left[\widehat{\boldsymbol{V}}_{R-1}^T, \quad \mathbf{1}_R - \widehat{\boldsymbol{V}}_{R-1}^T\mathbf{1}_{R-1}\right] \\ &= \widehat{\mathbf{A}}\widehat{\boldsymbol{V}}_R^T\end{aligned} \tag{27}$$

where $\widehat{\boldsymbol{V}}_R = [\widehat{\boldsymbol{V}}_{R-1}^T, \mathbf{1}_R - \widehat{\boldsymbol{V}}_{R-1}^T\mathbf{1}_{R-1}]^T$ is an $R \times R$ matrix. The final abundance estimation procedure, including the LL-GPLVM presented in Section III and the scaling procedure investigated in this section is referred to as fully constrained LL-GPVLM (FCLL-GPLVM) (a detailed algorithm is available in [22]). Once the final abundance matrix $\widehat{\mathbf{A}}$ and the matrix $\widehat{\boldsymbol{V}}_R$ have been estimated, we propose an endmember extraction procedure based on GP regression. This method is discussed in the next section.

## V. Gaussian Process Regression

Endmember estimation is one of the main issues in SU. Most of the existing EEAs intend to estimate the endmembers from the data, i.e., selecting the most pure pixels in the observed image [30]–[32]. However, these approaches can be inefficient when the image does not contain enough pure pixels. Some other EEAs based on the minimization of the volume containing the data (such as the minimum volume simplex analysis [33]) can mitigate the absence of pure pixels in the image. This section studies a new endmember estimation strategy based

on GP regression for nonlinear mixtures. This strategy can be used even when the scene does not contain pure pixels. It assumes that all the image abundances have been estimated using the algorithm described in Section IV. Consider the set of pixels $\{\mathbf{y}(n)\}_{n=1,\ldots,N}$ and the corresponding estimated abundance vectors $\{\hat{\boldsymbol{a}}(n)\}_{n=1,\ldots,N}$. GP regression first allows the nonlinear mapping $\boldsymbol{g}(\cdot)$ in (1) (from the abundance space to the observation space) to be estimated. The estimated mapping is denoted as $\hat{\boldsymbol{g}}(\cdot)$. Then, it is possible to use the prediction capacity of GPs to predict the spectrum $\hat{\boldsymbol{g}}(\boldsymbol{a})$ corresponding to any new abundance vector $\boldsymbol{a}$. In particular, the predicted spectra associated with pure pixels, i.e., the endmembers, correspond to abundance vectors that are the vertices of the simplex defined by (4). This section provides more details about GP prediction for endmember estimation. It can be seen from the marginalized likelihood (20) that $f(\mathbf{Y}|\boldsymbol{X}, \overline{\boldsymbol{P}}, \boldsymbol{U}, s^2, \sigma^2)$ is the product of $L$ independent GPs associated with each spectral band of the data space (21). Looking carefully at the covariance matrix of $\mathbf{y}_\ell$ (i.e., to $\boldsymbol{\Sigma} = s^2\boldsymbol{C}\boldsymbol{C}^T + \sigma^2\mathbf{I}_N$), we can write

$$\mathbf{y}_\ell = \boldsymbol{z}_\ell + \boldsymbol{\epsilon}_\ell \tag{28}$$

where $\boldsymbol{\epsilon}_\ell$ is the $N \times 1$ white Gaussian noise vector associated with the $\ell$th spectral band (having covariance matrix $\sigma^2\mathbf{I}_N$) and[2]

$$\boldsymbol{z}_\ell \sim \mathcal{N}(\boldsymbol{z}_\ell|\boldsymbol{\Psi}_x\boldsymbol{U}\bar{\boldsymbol{p}}_\ell, \boldsymbol{K}) \tag{29}$$

with $\boldsymbol{K} = s^2\boldsymbol{\Psi}_x\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Psi}_x^T$ the $N \times N$ covariance matrix of $\boldsymbol{z}_\ell$. The $N \times 1$ vector $\boldsymbol{z}_\ell$ is referred to as hidden vector associated with the observation $\mathbf{y}_\ell$. Consider now an $L \times 1$ test data with hidden vector $\boldsymbol{z}^* = [z_1^*, \ldots, z_L^*]^T$, abundance vector $\boldsymbol{a}^* = [z_1^*, \ldots, z_L^*]^T$ and $\boldsymbol{\psi}_x^* = \boldsymbol{\psi}[\boldsymbol{V}_R\boldsymbol{a}^*]$. We assume that the test data share the same statistical properties as the training data $\mathbf{y}_1, \ldots, \mathbf{y}_L$ in the sense that $[\boldsymbol{z}_\ell^T, z_\ell^*]$ is a Gaussian vector such that

$$\begin{bmatrix}\boldsymbol{z}_\ell \\ z_\ell^*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{z}_\ell \\ z_\ell^*\end{bmatrix} \middle| \begin{bmatrix}\boldsymbol{\Psi}_x\boldsymbol{U}\bar{\boldsymbol{p}}_\ell \\ \boldsymbol{\psi}_x^{*T}\boldsymbol{U}\bar{\boldsymbol{p}}_\ell\end{bmatrix}, \begin{bmatrix}\boldsymbol{K} & \boldsymbol{\kappa}(\boldsymbol{a}^*) \\ \boldsymbol{\kappa}(\boldsymbol{a}^*)^T & \sigma_{\boldsymbol{a}^*}^2\end{bmatrix}\right) \tag{30}$$

where $\sigma_{\boldsymbol{a}^*}^2 = s^2\boldsymbol{\psi}_x^{*T}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\psi}_x^*$ is the variance of $z_\ell^*$ and $\boldsymbol{\kappa}(\boldsymbol{a}^*)$ contains the covariances between the training inputs and the test inputs, i.e.,

$$\boldsymbol{\kappa}(\boldsymbol{a}^*) = s^2\boldsymbol{\psi}_x^{*T}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Psi}_x. \tag{31}$$

Straightforward computations leads to

$$z_\ell^*|\mathbf{y}_\ell \sim \mathcal{N}\left(z_\ell^*|\mu_\ell, s_l^2\right) \tag{32}$$

with

$$\begin{aligned}\mu_\ell &= \boldsymbol{\psi}_x^{*T}\boldsymbol{U}\bar{\boldsymbol{p}}_\ell + \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\boldsymbol{K} + \sigma^2\mathbf{I}_N)^{-1}(\mathbf{y}_\ell - \boldsymbol{\Psi}_x\boldsymbol{U}\hat{\boldsymbol{p}}_\ell) \\ s_l^2 &= \sigma_{\boldsymbol{a}^*}^2 - \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\boldsymbol{K} + \sigma^2\mathbf{I}_N)^{-1}\boldsymbol{\kappa}(\boldsymbol{a}^*).\end{aligned}$$

Since the posterior distribution (32) is Gaussian, the MAP and MMSE estimators of $\boldsymbol{z}^*$ equal the posterior mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_L)^T$.

In order to estimate the endmembers, we propose to replace the parameters $\boldsymbol{X}$, $\boldsymbol{U}$, $s^2$ and $\sigma^2$ by their estimates $\widehat{\boldsymbol{X}}^{(c)}$, $\widehat{\boldsymbol{U}}$, $\hat{s}^2$ and $\hat{\sigma}^2$ and to compute the estimated hidden vectors associated with the abundance vectors $\boldsymbol{a}^* = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$ for $r = 1, \ldots, R$. For each value of $r$, the $r$th estimated hidden vector will be the $r$th estimated endmember.[3] Indeed, for the LMM and the bilinear models considered in this paper, the endmembers are obtained by setting $\boldsymbol{a} = \boldsymbol{a}^* = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$ in the model (2) relating the observations to the abundances. Note that the proposed endmember estimation procedure provides the posterior distribution of each endmember via (32) which can be used to derive confidence intervals for the estimates. The next section presents some simulation results obtained for synthetic and real data.

[2]Note that all known conditional parameters have been omitted for brevity.

TABLE I
AREs: SYNTHETIC IMAGES

| | ARE ($\times 10^{-2}$) | | | | | |
|---|---|---|---|---|---|---|
| | $I_1$ | $I_2$ | $I_3$ | $I_1^*$ | $I_2^*$ | $I_3^*$ |
| PCA | **0.99** | 1.08 | 1.04 | **1.00** | 1.06 | 1.03 |
| LL-GPLVM | **0.99** | **0.99** | **1.00** | **1.00** | **1.00** | **0.99** |
| SU | 1.00 | 1.13 | 1.06 | 1.14 | 1.57 | 1.12 |
| FCLL-GPLVM | **0.99** | **0.99** | **1.00** | **1.00** | **1.00** | **0.99** |

## VI. SIMULATIONS ON SYNTHETIC DATA

### A. Subspace Identification

The performance of the proposed GPLVM for dimensionality reduction is first evaluated on three synthetic images of $N = 2500$ pixels. The $R = 3$ endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software [34] (i.e., green grass, olive green paint and galvanized steel metal). Additional simulations conducted with different endmembers are available in [22]. The first image $I_1$ has been generated according to the linear mixing model (LMM). The second image $I_2$ is distributed according to the bilinear mixing model introduced in [5], referred to as the "Fan model" (FM). The third image $I_3$ has been generated according to the generalized bilinear model (GBM) studied in [9] with the following nonlinearity parameters

$$\gamma_{1,2} = 0.9, \quad \gamma_{1,3} = 0.5, \quad \gamma_{2,3} = 0.3.$$

The abundance vectors $\boldsymbol{a}_n$, $n = 1, \ldots, N$ have been randomly generated according to a uniform distribution on the admissible set defined by the positivity and sum-to-one constraints (4). The noise variance has been fixed to $\sigma^2 = 10^{-4}$, which corresponds to a signal-to-noise ratio $\mathrm{SNR} \approx 30$ dB which corresponds to the worst case for current spectrometers. The hyperparameter $\gamma$ of the latent variable prior (15) has been fixed to $\gamma = 10^3$ and the number of neighbors for the LLE is $K = R$ for all the results presented in this paper. The quality of dimensionality reduction of the GPLVM can be measured by the average reconstruction error (ARE) defined as

$$\mathrm{ARE} = \sqrt{\frac{1}{LP}\sum_{n=1}^{N}\|\hat{\mathbf{y}}_n - \mathbf{y}n\|^2} \tag{33}$$

where $\mathbf{y}_n$ is the $n$th observed pixel and $\hat{\mathbf{y}}_n$ its estimate. For the LL-GPLVM, the $n$th estimated pixel is given by $\hat{\mathbf{y}}_n = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{U}}^T\boldsymbol{\psi}[\hat{\boldsymbol{x}}(n)]$ where $\widehat{\boldsymbol{P}}$ is estimated using (24). Table I compares

[3]Note that the estimated endmembers are centered since the data have previously been centered. The actual endmembers can be obtained by adding the empirical mean to the estimated endmembers.

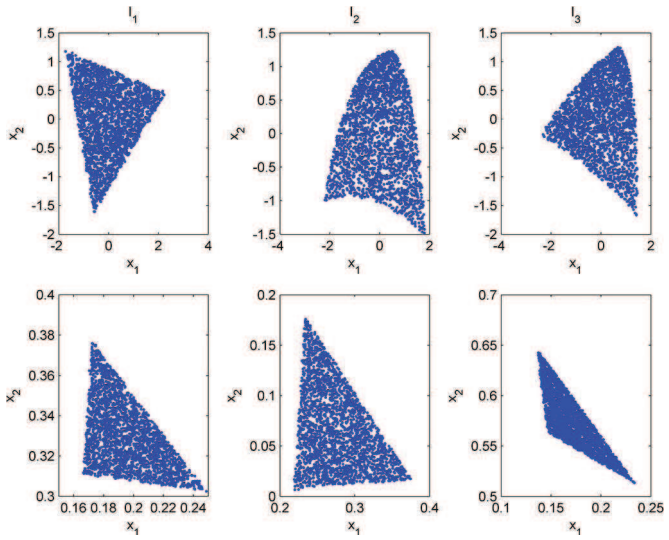


Fig. 4. Top: Representation of the $N = 2500$ pixels (dots) using the first two principal components provided by the standard PCA for the three synthetic images $I_1$ to $I_3$. Bottom: Representation using the latent variables estimated by the LL-GPLVM for the three synthetic images $I_1$ to $I_3$.

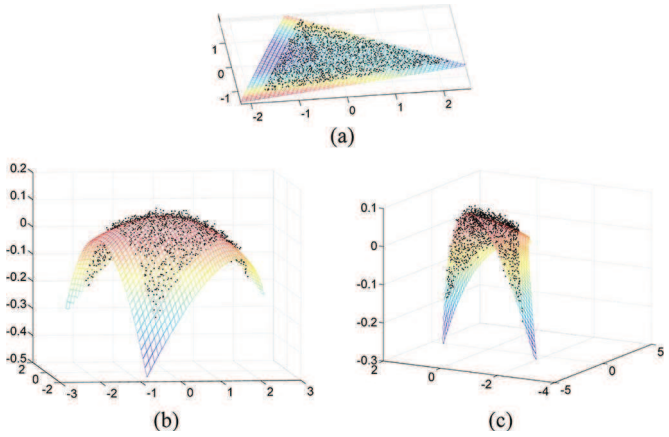



Fig. 5. Manifolds spanned by the $N = 2500$ pixels (black dots) of $I_1$, $I_2$ and $I_3$ using the 3 most significant PCA axes. The colored surface is the manifold identified by the LL-GPLVM.

the AREs obtained by the proposed LL-GPLVM and the projection onto the first $(R-1)$ principal vectors provided by the PCA. The proposed LL-GPLVM slightly outperforms PCA for nonlinear mixtures in term of ARE. More precisely, the AREs of the LL-GPLVM mainly consist of the noise errors ($\sigma^2 = 10^{-4}$), whereas model errors are added when applying PCA to nonlinear mixtures. Fig. 4 compares the latent variables obtained after maximization of (20) for the three images $I_1$ to $I_3$ with the projections obtained by projecting the data onto the $R-1$ principal vectors provided by PCA. Note that only $R-1$ dimensions are needed to represent the latent variables (because of the sum-to-one constraint). From this figure, it can be seen that the latent variables of the LL-GPLVM describe a noisy simplex for the three images. It is not the case when using PCA for the nonlinear images. Fig. 5 shows the manifolds estimated by the LL-GPLVM for the three images $I_1$ to $I_3$. This figure shows that the proposed LL-GPLVM can model the manifolds associated with the image pixels with good accuracy.

TABLE II
RNMSEs: SYNTHETIC IMAGES

| | RNMSE ($\times 10^{-3}$) | | | | | |
|---|---|---|---|---|---|---|
| | $I_1$ | $I_2$ | $I_3$ | $I_1^*$ | $I_2^*$ | $I_3^*$ |
| SU | 5.7 | 7.4 | 22.7 | 49.3 | 86.6 | 47.8 |
| FCLL-GPLVM | **3.9** | **4.2** | **5.4** | **4.8** | **7.2** | **7.5** |

### B. Abundance and Endmember Estimation

The quality of SU can be evaluated by comparing the estimated and actual abundances using the root normalized mean square error (RNMSE) defined by

$$\text{RNMSE} = \sqrt{\frac{1}{NR}\sum_{n=1}^{N}\|\hat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|^2} \tag{34}$$

where $\boldsymbol{a}_n$ is the $n$th actual abundance vector and $\hat{\boldsymbol{a}}_n$ its estimate. Table II compares the RNMSEs obtained with different unmixing strategies. The endmembers have been estimated by the VCA algorithm in all simulations. The algorithms used for abundance estimation are the FCLS algorithm proposed in [35] for $I_1$, the LS method proposed in [5] for $I_2$ and the gradient-based method proposed in [9] for $I_3$. These procedures are referred to as "SU" in the table. These strategies are compared with the proposed FCLL-GPLVM. As mentioned above, the Bayesian algorithm for joint estimation of $\mathbf{A}$ and $\boldsymbol{V}$ under positivity and sum-to-one constraints for $\mathbf{A}$ (introduced in [29]) is used in this paper for the scaling step. It can be seen that the proposed FCLL-GPLVM is general enough to accurately approximate the considered mixing models since it provides the best results in term of abundance estimation.

The quality of reconstruction of the unmixing procedure is also evaluated by the ARE. For the FCLL-GPLVM, the $n$th reconstructed pixel $\hat{\mathbf{y}}_n$ is given by $\hat{\mathbf{y}}_n = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{U}}^T\boldsymbol{\psi}[\hat{\boldsymbol{x}}^{(c)}(n)]$. Table I shows the AREs corresponding to the different unmixing strategies. The proposed FCLL-GPLVM outperforms the other strategies in term of ARE for these images.

Finally, the performance of the FCLL-GPLVM for endmember estimation is evaluated by comparing the estimated endmembers with the actual spectra. The quality of endmember estimation is evaluated by the spectral angle mapper (SAM) defined as

$$\text{SAM} = \arccos\left(\frac{\langle\hat{\mathbf{m}}_r, \mathbf{m}_r\rangle}{\|\hat{\mathbf{m}}_r\|\|\mathbf{m}_r\|}\right) \tag{35}$$

where $\mathbf{m}_r$ is the $r$th actual endmember and $\hat{\mathbf{m}}_r$ its estimate. Table III compares the SAMs obtained for each endmember using the VCA algorithm, the nonlinear EEA presented in [18] (referred to as "Heylen") and the FCLL-GPLVM for the three images $I_1$ to $I_3$. These results show that the FCLL-GPLVM provides accurate endmember estimates for both linear and nonlinear mixtures.

### C. Performance in Absence of Pure Pixels

The performance of the proposed unmixing algorithm is also tested in scenarios where pure pixels are not present in the observed scene. More precisely, the simulation parameters remain the same for the three images $I_1$ to $I_3$ except for the $N = 2500$

TABLE III
SAMs ($\times 10^{-2}$): SYNTHETIC IMAGES

| | | VCA | Heylen | FCLL-GPLVM |
|---|---|---|---|---|
| $I_1$ | $\mathbf{m}_1$ | **0.43** | 1.94 | 0.52 |
| | $\mathbf{m}_2$ | **0.22** | 0.66 | 0.86 |
| | $\mathbf{m}_3$ | 0.22 | 0.78 | **0.15** |
| $I_2$ | $\mathbf{m}_1$ | 1.62 | 0.75 | **0.33** |
| | $\mathbf{m}_2$ | 2.08 | 1.69 | **0.53** |
| | $\mathbf{m}_3$ | 1.15 | 0.42 | **0.34** |
| $I_3$ | $\mathbf{m}_1$ | 1.91 | 1.80 | **0.44** |
| | $\mathbf{m}_2$ | 1.36 | 0.86 | **0.58** |
| | $\mathbf{m}_3$ | 0.88 | 1.38 | **0.30** |

TABLE IV
SAMs ($\times 10^{-2}$): SYNTHETIC IMAGES

| | | VCA | Heylen | FCLL-GPLVM |
|---|---|---|---|---|
| $I_1^*$ | $\mathbf{m}_1$ | 2.87 | 6.38 | **0.38** |
| | $\mathbf{m}_2$ | 2.15 | 11.11 | **1.30** |
| | $\mathbf{m}_3$ | 2.10 | 2.62 | **0.24** |
| $I_2^*$ | $\mathbf{m}_1$ | 5.22 | 7.53 | **0.67** |
| | $\mathbf{m}_2$ | 8.02 | 9.59 | **1.46** |
| | $\mathbf{m}_3$ | 7.10 | 2.48 | **0.53** |
| $I_3^*$ | $\mathbf{m}_1$ | 6.89 | 6.59 | **0.61** |
| | $\mathbf{m}_2$ | 6.03 | 5.95 | **1.75** |
| | $\mathbf{m}_3$ | 3.73 | 2.36 | **0.48** |

abundance vectors, that are drawn from a uniform distribution in the following set

$$\left\{ \boldsymbol{a} \middle| \sum_{r=1}^{R} a_r = 1, 0.9 \geq a_r(n) \geq 0, \forall r \in \{1, \ldots, R\} \right\}. \quad (36)$$

The three resulting images are denoted as $I_1^*$, $I_2^*$ and $I_3^*$. Table I shows that the absence of pure pixels does not change the AREs significantly when they are compared with those obtained with the images $I_1$ to $I_3$. Moreover, FCLL-GPLVM is more robust to the absence of pure pixels than the different SU methods. The good performance of FCLL-GPVLM is due in part to the scaling procedure. Table II shows that the performance of the FCLL-GPLVM in term of RNMSE is not degraded significantly when there is no pure pixel in the image contrary to the situation where when the endmembers are estimated using VCA. Table IV shows the performance of the FCLL-GPLVM for endmember estimation when there is no pure pixel in the image. The results of the FCLL-GPLVM do not change significantly when they are compared with those obtained with images $I_1$ to $I_3$, which is not the case for the two other EEAs. The accuracy of the endmember estimation is illustrated in Fig. 6 which compares the endmembers estimated by the FCLL-GPLVM (blue lines) to the actual endmember (red dots) and the VCA estimates (black line) for the image $I_2^*$.

### D. Performance With Respect to Endmember Variability

The proposed method assumes that the spectrum characterizing a given material (i.e., an endmember) is unique for all the image pixels. This assumption has been widely used in linear unmixing, which has motivated the consideration of unique endmembers in this paper. However, taking endmember variability into consideration is also an important problem, depending on the observation conditions and the observed scene [36]–[38]. To

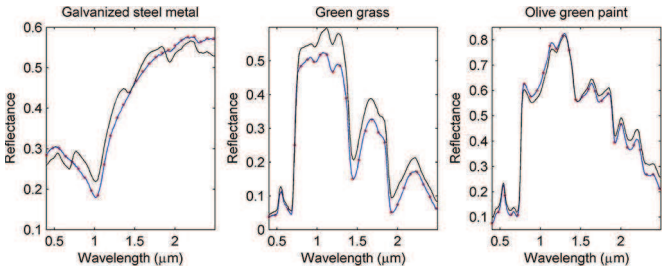



Fig. 6. Actual endmembers (red dots) and endmembers estimated by the FCLL-GPLVM (blue lines) and VCA (black line) for the image $I_2^*$.

evaluate the robustness of the proposed method to endmember variability, additional experiments have been performed. More precisely, sets of $N = 2500$ synthetic pixels have been generated according to the following nonlinear model

$$\mathbf{y}(n) = \sum_{r=1}^{R} \mathbf{m}_r(n) a_r(n) + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_i(n) a_j(n) \mathbf{m}_i(n) \odot \mathbf{m}_j(n) + \mathbf{e}(n),$$

where $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ has been generated uniformly in the simplex defined by the positivity and sum-to-one constraints and endmember variability has been considered using random endmembers, i.e., $\mathbf{m}_r(n) \sim \mathcal{N}(\mathbf{m}_r^0, \sigma_{\mathrm{M}}^2 \mathbf{I})$ where $\mathbf{m}_r^0$, $r = 1, \ldots, R$ are the actual endmembers extracted from the spectral library and $\sigma_{\mathrm{M}}^2$ is the endmember variance. Note that this model is similar to the Fan model studied in [5] except that the endmembers are random. Table V compares the performance of the proposed method with the performance of an unmixing strategy based on VCA (for endmember extraction) and the least squares method of [5] (for abundance estimation). This procedure is referred to as "SU" in the table. Four values of $\sigma_{\mathrm{M}}^2$ have been considered. The higher $\sigma_{\mathrm{M}}^2$, the higher the endmember variability. For each row, the best result has been highlighted in blue. The spectral angle mappers (SAMs) presented in Table V represent the angles between the estimated endmembers and the actual endmembers $\mathbf{m}_r^0$, $r = 1, \ldots, R$. From this table, it can be seen that for each value of $\sigma_{\mathrm{M}}^2$, the proposed method provides more accurate abundance and endmember estimates (in term of RNMSE and SAM, respectively), when compared with the SU approach. In particular, the performance of the proposed method is not significantly degraded for weak endmember variability.

## VII. APPLICATION TO A REAL DATASET

The real image considered in this section was acquired in 2010 by the Hyspex hyperspectral scanner over Villelongue, France (00°03′W and 42°57′N). $L = 160$ spectral bands were recorded from the visible to near infrared with a spatial resolution of 0.5 m. This dataset has already been studied in [39] and is composed of a forested area containing 12 identified vegetation species (ash tree, oak tree, hazel tree, locust tree, chestnut tree, lime tree, maple tree, beech tree, birch tree, willow tree, walnut tree and fern). More details about the data acquisition and pre-processing steps are available in [39]. The sub-image of size 50 × 50 pixels chosen here to evaluate the proposed unmixing procedure is depicted in Fig. 7. A reasonably small

TABLE V
ENDMEMBER VARIABILITY: SYNTHETIC IMAGES

| | | | $\sigma^2_{\mathrm{M}} = 0$ | $\sigma^2_{\mathrm{M}} = 10^{-5}$ | $\sigma^2_{\mathrm{M}} = 10^{-4}$ | $\sigma^2_{\mathrm{M}} = 10^{-3}$ |
|---|---|---|---|---|---|---|
| RNMSE ($\times 10^{-3}$) | SU | | 7.4 | 9.1 | 11.0 | 23.2 |
| | FCLL-GPLVM | | **4.2** | **5.1** | **9.1** | **11.9** |
| SAM ($\times 10^{-2}$) | SU | $\mathbf{m}_1$ | 1.62 | 2.06 | 2.09 | 2.55 |
| | | $\mathbf{m}_2$ | 2.08 | 1.67 | **1.94** | 2.38 |
| | | $\mathbf{m}_3$ | 1.15 | 1.27 | 1.06 | 1.23 |
| | FCLL-GPLVM | $\mathbf{m}_1$ | **0.33** | **0.44** | **0.46** | **1.59** |
| | | $\mathbf{m}_2$ | **0.53** | **0.71** | 2.01 | **1.62** |
| | | $\mathbf{m}_3$ | **0.34** | **0.47** | **0.42** | **0.99** |

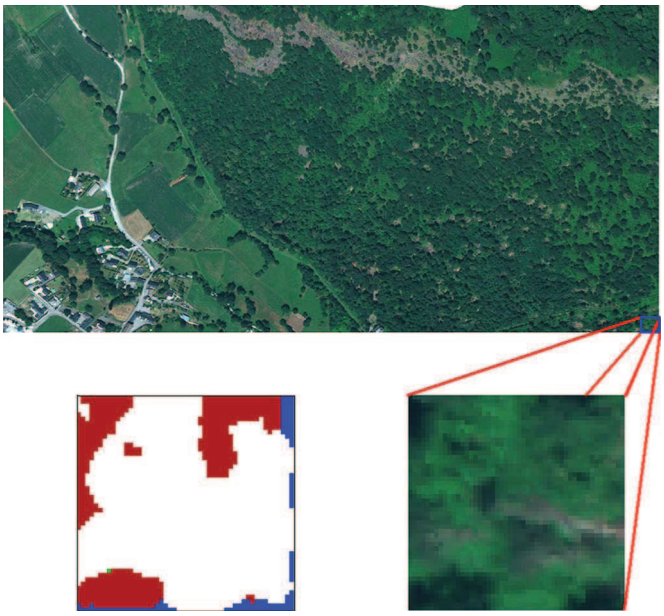


Fig. 7. Top: real hyperspectral Madonna data acquired by the Hyspex hyperspectral scanner over Villelongue, France. Bottom right: Region of interest shown in true colors (right). Bottom left: Classification map obtained in [39] for the region of interest. The labeled pixels are classified as Oak tree (red), Chestnut tree (blue), Ash tree (green) and non-planted-tree pixels (white).

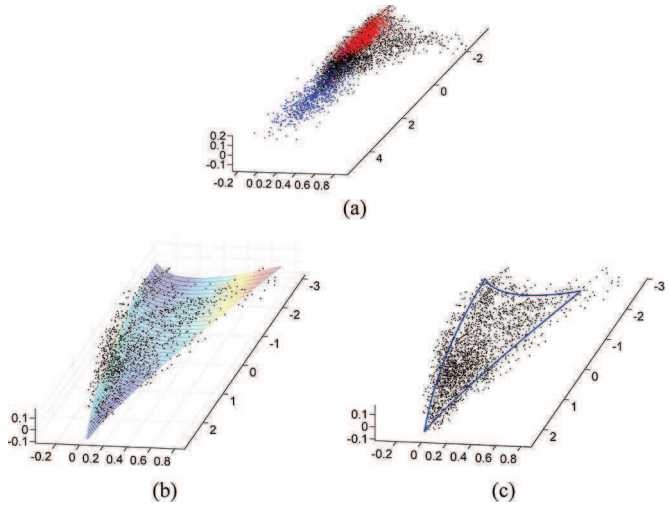



Fig. 8. (a) Representation of the $N = 2500$ pixels (black dots) of the Madonna image and the reference clusters corresponding to oak trees (red dots) and chestnut trees (blue dots) using the first three principal components provided by the standard PCA. (b) Representation of the $N = 2500$ pixels (dots) of the Madonna data and manifold identified by the LL-GPLVM (colored surface). (c) Representation of the $N = 2500$ pixels (dots) of the Madonna data and boundaries of the estimated transformed simplex (blue lines).

image is considered in this paper to ease the explanation of the results and to keep the processing overhead quite low. This image contains vegetation species of varying spatial density such that some pixels do not contain identified tree species. More precisely, the scene is mainly composed of three components since the data belong to a two-dimensional manifold (see black dots of Fig. 8(a)). Consequently, we assume that the scene is composed of $R = 3$ endmembers.[4] We propose to use the set of 32224 label spectra used in [39] for the learning step of the classification method presented herein to identify the components present in the area of interest. More precisely, Fig. 8(a) shows the reference clusters corresponding to oak trees (red dots) and chestnut trees (blue dots) projected in a 3-dimensional subspace (defined by the first three principal components of a PCA applied to the image of Fig. 7). These two clusters are the two closest sets of pixels to vertices of the data cloud. Consequently, oak and chestnut trees are identified as endmembers present in the image. Moreover, the new identified endmember is associated with the non-vegetation area (the strategy conducted in [39] was restricted to vegetation species). In the sequel, this endmember will be referred to as Endmember ♯3.

The simulation parameters have been fixed to $\gamma = 10^3$ and $K = R$. The latent variables obtained by maximizing the marginalized posterior distribution (10) are depicted in Fig. 9 (blue dots). It can be seen from this figure that the latent variables seem to describe a noisy simplex. Fig. 8(b) shows the manifold estimated by the proposed LL-GPLVM. This figure illustrates the capacity of the LL-GPLVM for modeling the nonlinear manifold. Table VI (left) compares the AREs obtained by the proposed LL-GPLVM and the projection onto the first $R - 1 = 2$ principal vectors provided by PCA. The proposed LL-GPLVM slightly outperforms PCA for the real data of interest, which shows that the proposed nonlinear dimensionality reduction method is more accurate than PCA (linear dimensionality reduction) in representing the data. The scaling step presented in Section IV is then applied to the estimated latent variables. The estimated simplex defined by the latent variables is depicted in Fig. 9 (red lines). Fig. 8 (c) compares the boundaries of the estimated transformed simplex with the image pixels. The abundance maps obtained after the scaling step are shown in Fig. 10 (top). The results of the unmixing procedure using the FCLL-GPLVM are compared

[4]Results of simulations conducted for different values of $R$ have been omitted in this paper for brevity. The results are available in [22].

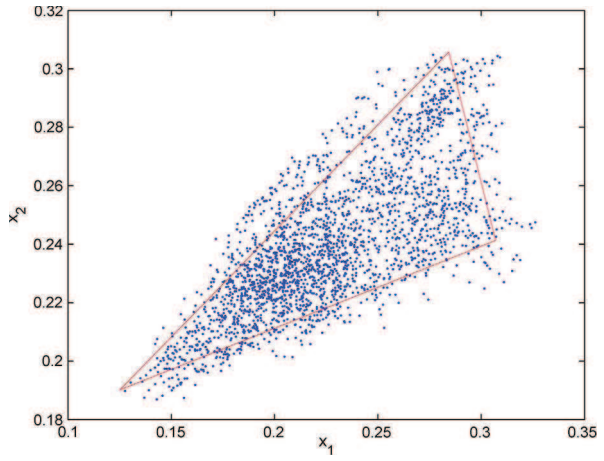



Fig. 9. Representation of the $N = 2500$ latent variables (dots) estimated by the LL-GPLVM and the simplex identified by the scaling step (red lines) for the Madonna data.

TABLE VI
AREs: Real Image ($\times 10^{-2}$)

| PCA | LL-GPLVM | VCA+FCLS | FCLL-GPLVM |
|---|---|---|---|
| 0.84 | **0.79** | 1.30 | **1.11** |

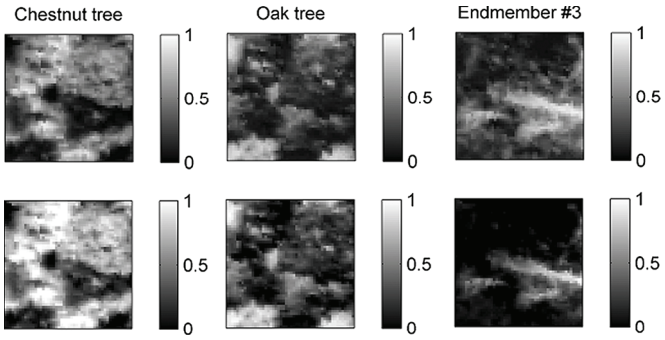



Fig. 10. Top: Abundance maps estimated using the FCLL-GPLVM for the Madonna image. Bottom: Abundance maps estimated using the VCA algorithm for endmember extraction and the FCLS algorithm for abundance estimation.

to an unmixing strategy assuming the LMM. More precisely, we use VCA to extract the endmembers from the data and use the FLCS algorithm for abundance estimation. The estimated abundance maps are depicted in Fig. 10 (bottom). The abundance maps obtained by the two methods are similar which shows the accuracy of the LMM as a first order approximation of the mixing model. However, the proposed unmixing strategy provides information about the nonlinearly mixed pixels in the image.

Moreover, Fig. 7 (bottom left) shows the classification map obtained in [39] for the region of interest. The white pixels correspond to areas where the classification method of [39] has not been performed. Since the aim of the work presented in [39] was to locate tree species, a non-planted-tree reference mask was used in [39] to classify only planted-tree pixels. Even if lots of pixels are not classified, the classified pixels can be compared with the estimated abundance maps. First, we can note the presence of the same tree species in the classification and abundance maps, i.e., oak and chestnut. We can also see that the pixels composed of chestnut trees and Endmember ♯3 are mainly located in the unclassified regions, which explains why they do not appear clearly in the classification map. Only one pixel is classified as being composed of ash trees in the region of interest. If unclassified pixels also contain ash trees, they are either too few or too mixed to be considered as mixtures of an additional endmember in the image.

Evaluating the performance of endmember estimation on real data is an interesting problem. However, comparison of the estimated endmembers with the ground truth is difficult here. First, since the nature of Endmember ♯3 is unknown, no ground truth is available for this endmember. Second, because of the variability of the ground truth spectra associated with each tree species, it is difficult to show whether VCA or the proposed FCLL-GPLVM provides the best endmember estimates. However, the AREs obtained for both methods (Table VI, right) show that the FCLL-GPLVM fits the data better than the linear SU strategy, which confirms the importance of the proposed algorithm for nonlinear spectral unmixing.

## VIII. Conclusions

We proposed a new algorithm for nonlinear spectral unmixing based on a Gaussian process latent variable model. The unmixing procedure assumed a nonlinear mapping from the abundance space to the observed pixels. It also considered the physical constraints for the abundance vectors. The abundance estimation was decomposed into two steps. Dimensionality reduction was first achieved using latent variables. A scaling procedure was then proposed to estimate the abundances. After estimating the abundance vectors of the image, a new endmember estimator based on Gaussian process regression was investigated. This decomposition of the unmixing procedure, consisting of first estimating the abundance vectors and subsequently the endmembers, breaks the usual paradigm of spectral unmixing. Simulations conducted on synthetic data illustrated the flexibility of the proposed model for linear and nonlinear spectral unmixing and provided promising results for abundance and endmember estimations even when there are few pure pixels in the image. It was shown in this paper that the proposed unmixing procedure provides better or comparable performance (in terms of abundance and endmember estimation) than state of the art unmixing strategies assuming specific mixing models. The choice of the nonlinear mapping used for the GP model is an important issue to ensure that the LL-GPLVM is general enough to handle different nonlinearities. In particular, different mappings could be used for intimate mixtures. In this paper, the number of endmembers was assumed to be known, which is not true in most practical applications. We think that estimating the number of components present in the image is an important issue that should be considered in future works. Finally, considering endmember variability in linear and nonlinear mixing models is an interesting prospect which is currently under investigation.

## Acknowledgment

The authors would like to thank Dr. M. Fauvel from the University of Toulouse—INP/ENSAT, Toulouse, France, for supplying the real image, reference data, and classification map related to the classification algorithm studied in [39] and used in this paper.

## REFERENCES


[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[3] J. Plaza, E. M. Hendrix, I. García, G. Martín, and A. Plaza, "On endmember identification in hyperspectral images without pure pixels: A comparison of algorithms," *J. Math. Imag. Vis.*, vol. 42, no. 2–3, pp. 163–175, Feb. 2012.

[4] B. Somers, M. Zortea, A. Plaza, and G. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012.

[5] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Remote Sens. Environ.*, vol. 30, no. 11, pp. 2951–2962, June 2009.

[6] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Physical modelling and non-linear unmixing method for urban hyperspectral images," in *Proc. IEEE GRSS Workshop on Hyperspectral Image Signal Process., Evolut. Remote Sens. (WHISPERS)*, Jun. 2011, pp. 1–4.

[7] B. Somers, K. Cools, S. Delalieux, J. Stuckens, D. V. der Zande, W. W. Verstraeten, and P. Coppin, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1183–1193, 2009.

[8] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE*, 2009, vol. 7477, pp. 74 770I–74 770I-8.

[9] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[10] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.

[11] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 39, no. 8, pp. 2314–2318, Aug. 2001.

[12] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2011, pp. 1151–1154.

[13] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, 2012.

[14] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2007, pp. 4041–4044.

[15] K.-H. Liu, E. Wong, and C.-I. Chang, "Kernel-based linear spectral mixture analysis for hyperspectral image classification," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evolut. Remote Sens. (WHISPERS)*, Aug. 2009, pp. 1–4.

[16] J. Chen, C. Richard, and P. Honeine, "A novel kernel-based nonlinear unmixing scheme of hyperspectral images," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2011, pp. 1898–1902.

[17] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, 2013.

[18] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 534–542, Jun. 2011.

[19] N. D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Proc. NIPS*, Vancouver, Canada, 2003.

[20] E. V. Bonilla, F. V. Agakov, and C. K. I. Williams, "Kernel multi-task learning using task-specific features," *J. Mach. Learning Res.- Proc. Tracks*, pp. 43–50, 2007.

[21] M. A. Alvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," Jun. 2011 [Online]. Available: http://arxiv.org/abs/1106.6251

[22] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," Univ. of Toulouse, Toulouse, France, Tech. Rep., Nov. 2012 [Online]. Available: http://altmann.perso.enseeiht.fr/

[23] J. Quionero-candela, C. E. Rasmussen, and R. Herbrich, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, p. 2005, 2005.

[24] N. D. Lawrence, "The Gaussian process latent variable model," Comput. Sci. Dept., Univ. of Sheffield, Tech. Rep., Jan. 2006 [Online]. Available: http://staffwww.dcs.shef.ac.uk/people/N.Lawrence/

[25] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 9, pp. 2601–2616, Sep. 2006.

[26] R. Urtasun, D. J. Fleet, and N. D. Lawrence, "Modeling human locomotion with topologically constrained latent variable models,," in *Proc. Conf. Human Motion, Understand., Model., Capture, Animat.*, Rio de Janeiro, Brazil, 2007, pp. 104–118.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2005.

[28] M. Brookes, *The Matrix Reference Manual*. London, U.K.: Imperial College, 2005.

[29] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 2657–2669, Nov. 2009.

[30] F. Chaudhry, C.-C. Wu, W. Liu, C.-I. Chang, and A. Plaza, "Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery," in *Recent Advances in Hyperspectral Signal and Image Processing*, C.-I. Chang, Ed. Trivandrum, Kerala, India: Research Signpost, 2006, ch. 2.

[31] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Vancouver, Canada, Apr. 1999, vol. 2, pp. 337–344.

[32] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[33] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Boston, MA, USA, Jul. 2008, vol. 3, pp. 250–253.

[34] *"ENVI User's Guide,"* ver. 4.0, RSI (Research Systems Inc.), Boulder, CO, USA, Sep. 2003, 80301.

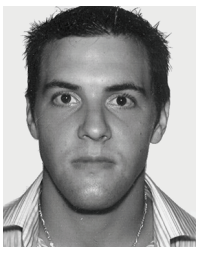
[35] D. C. Heinz and C.-I. Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 2001.

[36] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.

[37] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, 2011.

[38] A. Zare, P. Gader, T. Allgire, D. Drashnikov, and R. Close, "Bootstrapping for piece-wise convex endmember distribution detection," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process., Evolut. Remote Sens. (WHISPERS)*, May 2012, pp. 1–4.

[39] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, "Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2011, pp. 3672–3675.


**Yoann Altmann** (S'10) was born in Toulouse, France, in 1987. He received the Eng. degree in electrical engineering from the ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, France, both in 2010.

He is currently working towards the Ph.D. degree with the Signal and Communication Group, IRIT Laboratory, Toulouse, France.

**Nicolas Dobigeon** (S'05–M'08) was born in Angoulême, France, in 1981. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse (INP Toulouse), France, both in 2004, and the Ph.D. degree and Habilitation à Diriger des Recherches degree in signal processing from the INP Toulouse in 2007 and 2012, respectively.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. Since 2008, he has been with the National Polytechnic Institute of Toulouse (INP-ENSEEIHT, University of Toulouse), where he is currently an Associate Professor. He conducts his research within the Signal and Communications Group of the IRIT Laboratory, and he is also an affiliated faculty member of the Telecommunications for Space and Aeronautics (TeSA) cooperative laboratory. His recent research activities have been focused on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing, biomedical imaging, and genomics.

**Steve McLaughlin** (F'11) was born in Clydebank, Scotland, U.K., in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow, Scotland, U.K., in 1981 and the Ph.D. degree from the University of Edinburgh, Scotland, U.K., in 1990.

From 1981 to 1984, he was a Development Engineer in the industry, involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986, he worked on the design and development of high-frequency data communication systems. In 1986, he joined the Department of Electronics and Electrical Engineering at the University of Edinburgh as a Research Fellow, where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined the academic staff at Edinburgh, and from 1991 until 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he was awarded a personal Chair in Electronic Communication Systems at the University of Edinburgh. In October 2011, he joined Heriot-Watt University, Edinburgh, U.K., as a Professor of Signal Processing and Head of the School of Engineering and Physical Sciences. His research interests lie in the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy, and communication systems.

Prof. McLaughlin is a Fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology.

**Jean-Yves Tourneret** (SM'08) received the Ingénieur degree in electrical engineering from the École Nationale Supérieure d'Électronique, d Électrotechnique, dInformatique, dHydraulique, et des Télécommunications (ENSEEIHT), Toulouse, France, in 1989 and the Ph.D. degree from National Polytechnic Institute, Toulouse, France, in 1992.

He is currently a Professor with ENSEEIHT and a Member of the IRIT Laboratory (UMR 5505 of the CNRS). His current research interests include statistical signal processing with a particular interest to Bayesian methods and Markov chain Monte Carlo (MCMC) algorithms.

Prof. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), held in Toulouse, France, in 2002. He was also a member of the Organizing Committee for the International Conference of Acoustics, Speech and Signal Processing (ICASSP) in Toulouse, France, in 2006. He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society from 2001 to 2007 and 2010 to present. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011.